NAVAL HEALTH RESEARCH CENTER
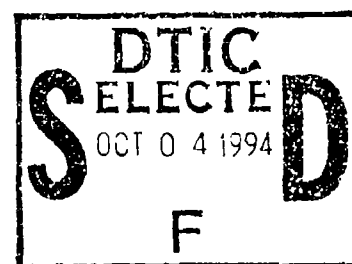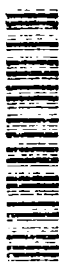
DEVELOPMENT OF AN EPIDEMIOLOGICAL DATABASE

MANAGEMENT, EXTRACTION, AND ANALYSIS

SYSTEM (EPISYS)

AD-A285 198

DTIC
SELECTED
OCT 0 4 1994
F

I. Show

M. R. White

94-31520

Technical Document 93-6F

DEVELOPMENT OF AN EPIDEMIOLOGICAL DATABASE MANAGEMENT, EXTRACTION, AND ANALYSIS SYSTEM (EPISYS)

Prepared for:

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
San Diego, CA 92186-5122

Prepared by:

Ivan Show, Ph.D.
SOUTHWEST RESEARCH ASSOCIATES INC.,
2006 Palomar Airport Road, Suite 207
Carlsbad, CA 92008

and

Martin R. White, M.P.H.
NAVAL HEALTH RESEARCH CENTER
San Diego, CA 92186-5122

Summary

Problem: Military health-care planners need to have the capability to easily detect, sample, and analyze any ICD9-CM illness or injury based upon a number of ancillary variables including age, race, sex, service branch, ship type, pay grade, and occupation if they are to provide the best health-care possible.

Objective: To provide military health-care planners with a integrated system of computer programs that will significantly improve the ability to access various medical files, allowing for the continued monitoring of the health and medical needs of U.S. military personnel.

Approach: A system of computer programs are being developed using the latest object-oriented technology designed to run on IBM-compatible PCs under the windows operating environment. The system Epidemiological Interactive System (EPISYS) consists of a number of user modules, epilimit, epibase, episam, epimips and a utilities module, designed to work as a integrated system. These modules each provide a useful function, and additional modules can be added as required.

Results: A system was developed to give researchers and medical planners the capability to easily detect, sample, and analyze any ICD9-CM illness or injury based upon a number of variables including age, race, sex, service branch, ship type, pay grade, and occupation. These programs will significantly improve the ability to access various medical files, and will provide investigators with an integrated system of computer programs for health monitoring and medical projection needs.

Conclusions: Providing medical researchers and military health-care planners with a system of computer programs to rapidly access medical information greatly assists them with making research, clinical, and management decisions. The work to integrate EPISYS, SAMS, and other medical systems will be studied in an effort to expand the utility of the system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

The goals of this project were to complete the specification and design of a computerized system to accomplish the following tasks:

- create methods to incorporate ICD9-CM incidence rate numerator and denominator data from all sources and automatically update analyses;

- specify a common format and framework for the maintenance of ICD9-CM incidence rate numerator and denominator data from all sources; and

- create methods for the surveillance of ICD9-CM incidence rates and for the detection and flagging of rates that exceed predefined thresholds.

Methods for incorporation of data from both inside and outside of the Naval Health Research Center (NHRC) was developed. The system design incorporates new data in a manner that makes it readily accessible to NHRC investigators and at the same time improves baseline incidence rate estimation. Data that can be handled includes historical data as well as data resulting from the sampling scenario developed under the previous Southwest Research Associates (SRA) contract. The system is designed to efficiently handle very large data sets, such as those maintained by the Argonne National Laboratory. System design allows automatic updating of sampling and analysis methods developed previously. System design also includes detecting and flagging ICD9-CM incidence rates that fall outside of predetermined threshold limits.

The system, when fully implemented, will result in a powerful databasing, analysis, and surveillance system. Based on inputs from NHRC investigators, the system ultimately puts NHRC personnel in closer touch with vast amounts of historical data and relieves them of the burden of manually extracting ICD9-CM rate data on a case-by-case basis.

During this project, specifications were developed and computer programs developed, tested, and validated. Incorporation of outside data and creation of databases was limited to testing and validating the computer programs.

An additional task was to modify the Medical Information Projection System (MIPS) so that it could be incorporated into the overall system design. This task involved improving the user interface and modifying the supporting file structures; the overall objective was partially accomplished during this project. Additional work remains to be completed.

2.0 SYSTEM OVERVIEW

The overall system includes four computer software systems plus a fifth semi-automated function. All five major system components are completely interconnected and integrated by means of user-friendly interfaces. The major components of EPISYS are (1) semiautomated data extraction, (2) Epibase, (3) Epilimit, (4) Episam, and (5) Epimips. The major data files at the heart of the system are the Common Data Format (CDF) files. Figure 1 shows the structure of the overall system (EPISYS) and the flow of data and information.

```
 ┌─────────────────┐       ┌─────────────────┐       ┌─────────────────┐
 │  Data Files and │       │  Semiautomated  │       │  Data Files and │
 │   Information   │ ────▶ │   Extraction    │ ◀──── │   Information   │
 │ External to NHRC│       │                 │       │Maintained by NHRC│
 └─────────────────┘       └────────┬────────┘       └─────────────────┘
                                    │
                                    ▼
                           ┌─────────────────┐
                           │Common Data Format│
                           │      Files       │
                           │ Maintained by NHRC│
                           └────────┬────────┘
                                    │
                                    ▼
                           ┌─────────────────┐
                           │     Epibase     │
                           │  User-Controlled │
                           │  Analytical Data │
                           │   Preparation   │
                           └─────────────────┘
   ┌─────────────────┐                           ┌─────────────────┐
   │    Epilimit     │ ◀──────────────────────▶ │     Episam      │
   │  User-Controlled │                          │  User-Controlled │
   │Threshold Detection│                          │  DNBI Sampling  │
   │                 │                          │   and Analysis  │
   └─────────────────┘                           └─────────────────┘
                           ┌─────────────────┐
                           │     Epimips     │
                           │  User-Controlled │
                           │   Projection and │
                           │ Analysis System │
                           └─────────────────┘
```

FIGURE 1. Structure and data flow of EPISYS. Lines and arrows indicate the flow of information among system components.

EPISYS is designed to be run from a single user-interface program and is controlled from an IBM-PC workstation. All of the major functions (except semiautomated data extraction) are performed on an IBM-PC workstation. The system performs the following tasks:

- accepts data from all sources, both inside and outside of NHRC;
- extracts and formats the data;
- performs surveillance for rates outside of set 'norms";
- performs sampling specification and multivariate analysis;
- performs rate projection analysis.

Semiautomated data extraction accepts new data into EPISYS and formats it for use by all other components of the system; this function and the CDF files are described in section 3.0 of this report. Epibase performs various database functions plus the actual data extraction; section 4.0 of this report describes Epibase. Epilimit performs surveillance of ICD9-CM incidence rates as new data enters the system or on demand; section 5.0 describes Epilimit. Episam performs various statistical analyses and specifies sampling requirements; it is described in a previous report to NHRC titled "Development of a Sampling Strategy for Disease and Non-Battle Injury (DNBI) Data Rates." Epimips projects expected incidence rates; a detailed description will appear in a later report to NHRC.

3.0 SEMIAUTOMATED DATA EXTRACTION AND CDF FILES

This function is the main entrance point for information into EPISYS. It can only be partially automated because each new data set accepted will be in a different format. For each individual case of new data input, adjustments to the format and logic of the input programs will be required. In all cases, however, output to the CDF files will be in the same format.

As currently performed for NHRC projects, data extraction is on a project-by-project basis and is extremely time-consuming, often taking days of computer time and weeks of real time. In addition, the same data extraction is sometimes repeated for different projects. Initial data extraction, as part of EPISYS, will still be time-consuming; not much can be done about that. However, extraction will only be performed once for each new data input and will always produce consistently formatted CDF files. The CDF files will then provide a rapid means of accessing data for individual projects. Common Data Format files will be in both ASCII and dBASE formats, allowing for the widest possible utility.

The Semiautomated Data Extraction process will be performed on the VAX system and will be entirely under the control of an EPISYS manager. Common Data Format files will be available on the VAX hard drives and will be archived on magnetic tapes. The general user will have "read-only" access to the CDF files through Epibase. Only the EPISYS manger will be able to modify or update the CDF files.

Table 1 gives the complete format for CDF rate denominator files. The file specification was carefully designed to contain all variables and information considered necessary for NHRC project use. The format is the same for ASCII, VAX, dBASE, and IBM-PC dBASE files.

TABLE 1. CDF File Format for Rate Numerator Data Files

1 - Social Security Number	Character	9
2 - Date of Birth	Integer	5
3 - Abbreviated Name	Character	10
4 - Sex	Integer	1
5 - Race	Integer	1
6 - Service Branch	Integer	1
7 - Pay Grade	Integer	2
8 - Rate Code	Integer	2
9 - Duty UIC	Character	5
10 - Platform	Integer	2
11 - Ocean Code	Integer	2
12 - Treatment Date	Integer	5
13 - Treatment UIC	Character	5
14 - Visit Number	Integer	1
15 - Return to Duty	Logical	1
16 - Diagnosis/Class 1	Integer	3,2
17 - Diagnosis/Class 2	Integer	3,2
18 - Diagnosis/Class 3	Integer	3,2
19 - Age	Integer	2

The CDF format is a system of two types of related files. The first file type, (variables 1 through 5 - Table 1), contains time-invariant personnel data. The other file type, (an index to the first file type plus variables 6 through 19 - Table 1), contains time-varying data. This indexed structure reduces data repetition and thus the size of the overall database. Most CDF file data is coded. Table 2 shows the essential structure of the codes and contains definitions for most of the variables. Tables 3 through 6 contain codes for the following variables: Major ICD9-CM Categories, Minor ICD9-CM Categories, Platforms, and Ocean Regions.

TABLE 2. Definitions for CDF Rate Numerator Data Files

1 - Social Security Number	-	9-digit number without "-"
2 - Date of Birth	-	modified Julian date (days since 01/01/1930)
3 - Abbreviated Name	-	first initial, middle initial, first 8 letters of last name
4 - Sex	-	coded (1,2) male, female
5 - Race	-	coded (1,2,3) white, black, other
6 - Service Branch	-	coded (1,2) USN, USMC
7 - Pay Grade	-	coded (E1, E2,.....,09)
8 - Rate Code	-	coded (1,2,.......,99)
9 - Duty UIC	-	5 character UIC or RUC
10 - Platform	-	coded (1,2,.......,47)
11 - Ocean Code	-	coded (1,2,......,14)
12 - Treatment Date	-	modified Julian date (days since 01/01/1930)
13 - Treatment UIC	-	5 character UIC or RUC
14 - Visit Number	-	visit sequence (1,2,....)
15 - Return to Duty	-	logical (Y/N)
16 - Diagnosis/Class 1	-	coded (5-digit ICD9-CM as modified by NHRC)
Diagnosis	-	coded (minor ICD9 category (1,2,........,130)
Class	-	coded (major ICD9 category (1,2,......29)
17 - Diagnosis/Class 2	-	secondary, coded (same as 16)
18 - Diagnosis/Class 3	-	tertiary, coded (same as 16)
19 - Age	-	age of patient at time of treatment

TABLE 3. Major ICD9-CM Codes and Categories

Code	Major ICD9-CM Category
1	Viral Diseases
2	Bacterial Diseases
3	Mycobacterial Diseases
4	Rickettsial Diseases
5	Chlamydial Diseases
6	Sexually Transmitted Diseases
7	Fungus Diseases
8	Spirochete-Caused Diseases
9	Protozoal Diseases
10	Ectoparasites
11	Worm-Caused Diseases
12	Diseases of Uncertain Etiology
13	Neoplasms
14	Endocrine/Nutrition/Metabolic Diseases
15	Diseases of Blood/Blood-Forming Organs
16	Mental Disorders
17	Diseases of Nervous System/Sense Organs
18	Diseases of Circulatory System
19	Diseases of Respiratory System
20	Diseases of Digestive System
21	Diseases of Genitourinary System
22	Conditions of Pregnancy/Childbirth
23	Diseases of Skin/Subcutaneous Tissue
24	Diseases of Musculoskeletal System
25	Congenital Anomalies
26	Perinatal Morbidity/Mortality
27	Symptomatic/Ill-Defined Conditions
28	Accident/Poisoning/Violence
29	Supplementary/Special Conditions

TABLE 4 Minor ICD9-CM Codes and Categories

Code	Major Code	Minor Category
1	2	Cholera
2	2	Typhoid Fever
3	2	Paratyphoid Fever
4	2	Other Salmonella Infections
5	2	Bacillary Dysentery
6	2	Bacillary Food Poisoning
7	9	Amebiasis
8	9	Other Protozoal Intestinal Diseases
10	2	Diarrheal Diseases
12	3	Tuberculosis of Pulmonary
13	3	Tuberculosis of Other Respiratory
14	3	Tuberculosis of Meninges , Central Nervous System
15	3	Tuberculosis of Intestine/Peritoneum/Mesenteric
16	3	Tuberculosis of Bones and Joints
17	3	Tuberculosis of Genitourinary System
18	3	Tuberculosis of Other Organs
19	3	Tuberculosis, Disseminated
20	3	Tuberculosis, Late Effects
21	3	Tuberculosis, Unspecifed Site
22	2	Plague
23	2	Tularemia
24	2	Anthrax
25	2	Brucellosis
26	2	Glanders
27	2	Melioidosis
28	2	Rat-Bite Fever
29	2	Other Zoonotic Bacterial Diseases
30	2	Leprosy
31	3	Other Mycobacterial Diseases
32	2	Diphtheria
33	2	Whooping Cough
34	2	Streptococcal Sore Throat/Scarlet Fever
35	2	Erysipelas
36	2	Meningococcal Infections
37	2	Tetanus
38	2	Septicemia
39	2	Other Bacterial Diseases
40	1	Acute Poliomyelitis
41	1	Aseptic Meningitis (Enterovirus)
42	1	Other Enteroviral Diseases of the Central Nervous System
43	1	Smallpox
44	1	Cowpox
45	1	Chickenpox
46	1	Herpes zoster
47	1	Herpes simplex
48	1	Measles
49	1	Rubella
50	1	Other Viral Exanthemata

TABLE 4. (continued)

Code	Major Code	Minor Category
51	1	Yellow Fever
52	1	Dengue
53	1	Mosquito-Borne Viral Encephalitis
54	1	Tick-Borne Viral Encephalitis
55	1	Unspecified Arthropod-Borne Encephalitis
56	1	Acute Encephalitis, Epidemic
57	1	Viral Encephalitis
58	1	Arthropod-Borne Hemorrhagic Fever
59	1	Epidemic Hemorrhagic Fever
60	1	Other Arthropod-Borne Diseases
61	1	Viral Hepatitis
62	1	Rabies
63	1	Mumps
64	4	Psittacosis
65	1	Coxsackie Virus Diseases
66	1	Infectious Mononucleosis
67	5	Trachoma
68	1	Other Viral Disease Conjunctiva
69	1	Other Viral Diseases
70	4	Louse-Borne Typhus
72	4	Tick-Borne Rickettsiosis
73	4	Other Rickettsiosis
74	9	Malaria
75	9	Leishmania
76	9	American Trypanosomiasis
77	9	Other Trypanosomiasis
78	8	Relapsing Fever
80	6	Syphilis, Congenital
81	6	Syphilis, Early Symptomatic
82	6	Syphillis, Early Latent
83	6	Syphillis, Cardiovascular
84	6	Syphillis, Central Nervous System
85	6	Syphillis, Other and Unspecified
86	6	Syphills
87	6	Gonococcal Infections
88	6	Other Venereal Diseases
89	8	Leptospirosis
90	8	Vincent's Angina
91	8	Yaws
92	8	Pinta
93	8	Other Spirochete Infections
94	7	Dermatophytosis
95	7	Dermatophytosis, Other and Unspecified
96	7	Moniliasis
97	7	Actinomycosis
98	7	Coccidiomycosis

TABLE 4 (continued)

Code	Major Code	Minor Category
100	7	Blastomycosis
101	7	Other Mycoses
102	11	Schistosomiasis
103	11	Other Trematode Infections
104	11	Hydatidosis
105	11	Other Cestode Infections
106	11	Trichiniasis
107	11	Filarial Infections
108	11	Ancylostomiasis
109	11	Other Intestinal Helminthiasis
110	11	Other and Unspecified Helminthiasis
111	11	Unspecified Intestinal Parasitism
112	9	Toxoplasmosis
113	6	Urogenital Trichomoniasis
114	10	Pediculosis
115	10	Acariasis
116	10	Other Infestations
117	10	Other Arthropod Infestations
118	12	Sarcoidosis
1002	13	Neoplasms
1003	14	Endocrine/Nutrition/Metabolic Diseases
1004	15	Diseases of Blood/Blood-Forming Organs
1005	16	Mental Disorders
1006	17	Diseases of Nervous System/Sense Organs
1007	18	Diseases of Circulatory System
1008	19	Diseases of Respiratory System
1009	20	Diseases of Digestive System
1010	21	Diseases of Genitourinary System
1011	22	Conditions of Pregnancy/Childbirth
1012	23	Diseases of Skin/Subcutaneous Tissue
1013	24	Diseases of Musculoskeletal System
1014	25	Congenital Anomalies
1015	26	Perinatal Morbidity/Mortality
1016	27	Symptomatic/Ill-Defined Conditions
1017	28	Accident/Poisoning/Violence
1018	29	Supplementary/Special Condition

TABLE 5. Platform Codes, Designations, and Definitions

Code	Designation	Platform/Ship Type
1	Ashore	Station Ashore
2	AE	Ammunition Ship
3	AF	Combat Stores Ship
4	AFS	Combat Stores Ship
5	AGF	Miscellaneous Command Ship
6	AGS	Survey Ship
7	AGSS	Survey Ship
8	AO	Oiler
9	AOE	Fast Combat Support Ship
10	AOR	Replenishment Oiler
11	AR	Repair Ship
12	ARL	Repair Ship
13	ARS	Salvage Ship
14	AS	Submarine Tender
15	ASR	Submarine Rescue Ship
16	ATF	Fleet Tug
17	BB	Battle Ship
18	CA	Cruiser
19	CG	Guided Missile Cruiser
20	CGN	Nuclear Guided Missile Cruiser
21	CV	Attack Aircraft Carrier
22	CVA	Assault Aircraft Carrier
23	CVS	Aircraft Carrier
24	DD	Destroyer
25	DDG	Guided Missile Destroyer
26	FF	Frigate
27	FFG	Guided Missile Frigate
28	LCC	Amphibious Command Ship
29	LHA	Amphibious Assault Ship
30	LHD	Amphibious Transport Dock
31	LKA	Amphibious Cargo Ship
32	LPD	Amphibious Transport Dock
33	LSD	Landing Ship Dock
34	LST	Landing Ship Tank
35	SS	Submarine
36	SSBN	Nuclear Ballistic Missile Submarine
37	SSN	Nuclear Attack Submarines

TABLE 6. Ocean Region Codes, Designations and Definitions

Code	Designation	Ocean Region
1	A1	Atlantic Ocean, Northwestern
2	A2	Arctic Ocean
3	A3	Caribbean Sea and Gulf of Mexico
4	A4	Atlantic Ocean, Southern
5	A5	Atlantic Ocean, Northeastern
6	A6	Mediterranean Sea
7	A7	Persian Gulf and Red Sea
8	P1	Pacific Ocean, Northeastern
9	P2	Pacific Ocean, Central
10	P3	Bering Sea
11	P4	Pacific Ocean, Western
12	P5	Pacific Ocean, Southeastern
13	P6	Indian Ocean
14	P7	Antarctic Ocean

4.0 EPIBASE: DATA SPECIFICATION AND EXTRACTION

The Epibase program is the heart of EPISYS. Epibase operates on the CDF files and extracts data in a format that is compatible with the rest of the system. Epibase produces both ASCII and dBASE files as specified by the user. Epibase consists of two main functions: (1) data structure definition and (2) data extraction. Data structure definition functions include defining the independent variable structure of the data to be extracted as well as saving and restoring, reviewing, and printing definitions. All functions are menu-driven.

The data structure definition functions are used to specify the independent variables for creation of CDF files. Data structure definitions are based on the following independent variables:

* major and minor ICD9-CM categories,
* sex,
* age group at time of treatment,
* service branch,
* pay grade at time of treatment,
* assignment platform (ashore and afloat) at time of treatment,
* location by oceanic region at time of treatment,
* rate/occupation code at time of treatment,
* treatment data.

The dependent variables extracted are ICD9-CM incidence rates. Rate data can be extracted as total number of incidences or can be normalized to predefined denominators.

The data extraction functions perform the actual data extraction from the CDF files and produce ASCII and dBASE files. Data extraction functions include actual data extraction, creation of new data files, and reviewing and printing data.

The following sections contain detailed descriptions of Epibase and can be used as a guide. Epibase is run from an IBM-PC workstation. Actual data extraction is performed on a VAX system. The extracted data is then usually transferred to the IBM-PC for further processing and analysis.

4.1 Basic Operation of the Program

Epibase is a completely menu-driven system. Three actions are required of the user; these are incorporated into three types of screens: (1) menu screens, (2) category selection screens, and (3) direct input screens. The following keys are used for all operations:

UP ARROW	move up one item; if at top, go to bottom
DOWN ARROW	move down one item; if at bottom, go to top
HOME	move to first item
END	move to last item
ENTER	select highlighted item
ESC	causes program to exit some screens

The most common operation is menu selection, there are two ways to select menu items: (1) press "ENTER" when a menu item is highlighted, and (2) press the menu item letter. Descriptive

messages appear at the bottom of all menu screens to aid in menu selection. When a menu item is highlighted, an appropriate message appears. When a menu item is selected in error, an error message appears. After reading the message, press any key to return to the menu.

Category selection screens are used to select multiple items from the list that appears on the screen. To select a category, highlight it and then press "ENTER". An asterisk will appear beside the selected item to indicate that it has been selected. The selection process can be repeated as many times as necessary. To "deselect" an item, highlight the item and then press "ENTER." The asterisk will disappear, indicating that the item is no longer selected. All category selection screens have a "Return to Previous Menu" item at the bottom of the selection list; when through making selections, highlight this item and press "ENTER."

Direct data entry screens require the user to type in requested information. The requirement for a direct data entry is indicated by the appearance of a query followed by a highlighted data entry block and then press "ENTER." Direct data entries are error checked to the extent possible. If nothing seems to happen when a value is entered, it probably means that the wrong data type was entered; for instance, a letter instead of a number. If a context error is made, an error message will return to the appropriate menu.

4.2 Entering and Saving a New Data Structure Definition

In this section, a data structure definition named "TEST1" is created step by step and then saved. When Epibase is invoked, the menu shown in Figure 2 appears. Select "A. Define Data Structure." The "Data Structure Control" menu (Figure 3) will appear. Select "A. Create/Edit Definition" and the "New/Edit Data Structure" menu (Figure 4) will appear.

```
             MAIN OPTIONS MENU
    ──────────────────────────────────────
    A.    Define Data Structure
    B.    Extract Data
    Q.    Exit Program
```
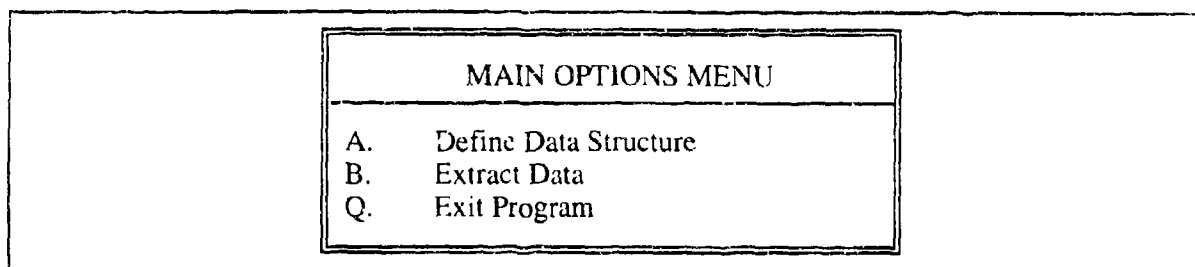
FIGURE 2. The Main Options Menu.

```
           DATA STRUCTURE CONTROL
    ──────────────────────────────────────
    A.    Create/Edit Definition
    B.    Select Existing Definition
    C.    Review Active Definition
    D.    Print Active Definition
    E.    Save Active Definition
    R.    Return to Main Options Menu
```
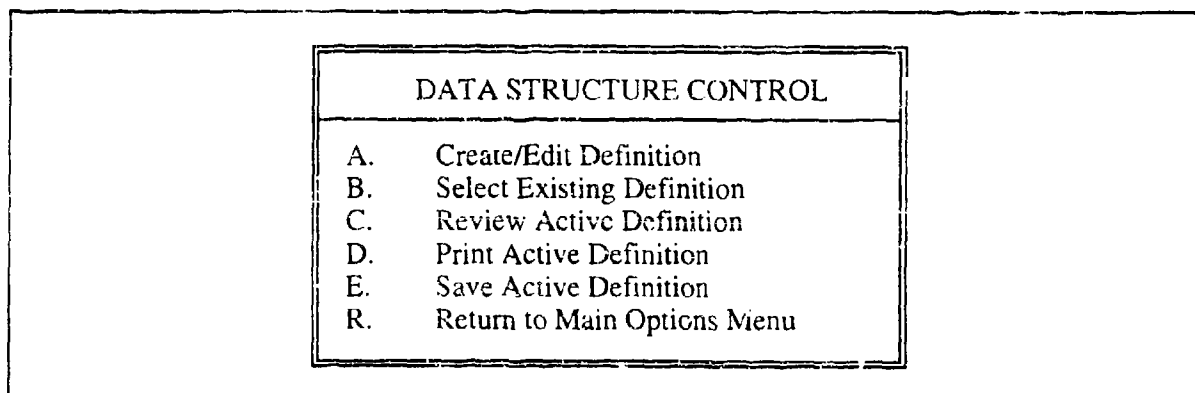
FIGURE 3. Data Structure Control Menu.

```
         NEW/EDIT DATA STRUCTURE
    ──────────────────────────────────────
    A.    ICD9-CM Categories
    B.    Sex Categories
    C.    Age Groups
    D.    Service Branch Categories
    E.    Platform Categories
    F.    Oceanic Region Categories
    G.    Rate/Occupation Code
    H.    Inclusive Dates
    I.    Calculation Parameters
    J.    Cancel All Changes
    R.    Return to Main Options Menu
```
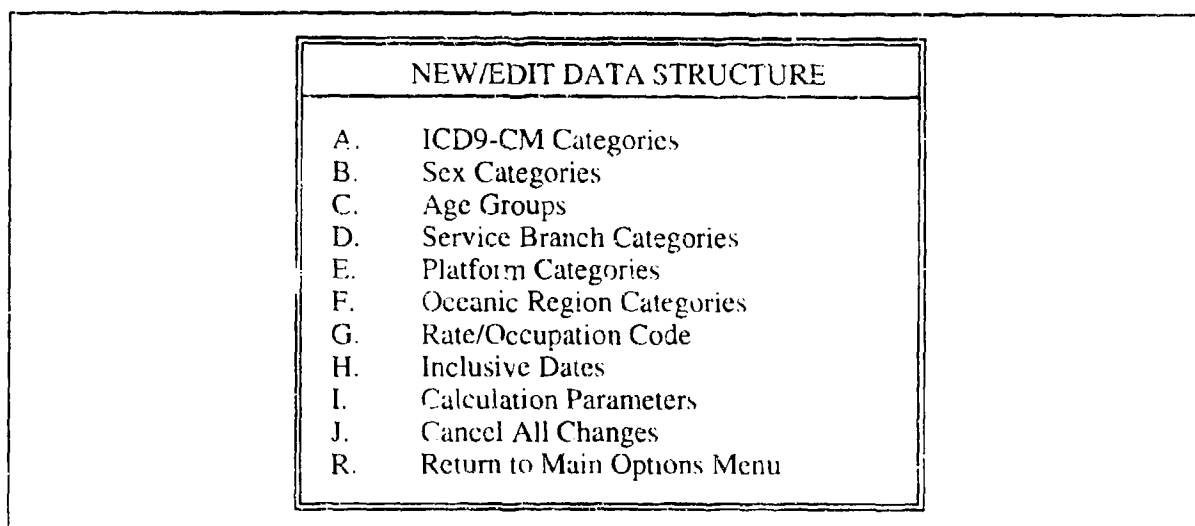
FIGURE 4. New/Edit Data Structure Menu.

To proceed with the data structure definition, select each of the items of the "New/Edit Data Structure" menu (Figure 3) in turn. Note that an item can be repeated as many times as necessary.

4.2.1 Specifying ICD9-CM Categories

Select "A. ICD9-CM Categories" (Figure 4). The "Data Definition ICD9-CM Categories" menu (Figure 5) will appear. Items A through C will automatically make the appropriate entry in the Data Structure Definition. Items D through F require the user to select specific ICD9-CM categories. Only one item can be selected. The item selected can, however, be changed at a later time. After an ICD9-CM selection is completed, the program will display a message to that effect and then return to the "New/Edit Data Structure" menu (Figure 4).

```
┌─────────────────────────────────────────────────────────────────┐
│                                                                   │
│        ┌──────────────────────────────────────────────┐          │
│        │  DATA DEFINITION - ICD9 CATEGORIES           │          │
│        ├──────────────────────────────────────────────┤          │
│        │  A.    Combine All Categories                │          │
│        │  B.    Specify All Major Categories          │          │
│        │  C.    Specify All Minor Categories          │          │
│        │  D.    Specify Individual Major Categories   │          │
│        │  E.    Specify Individual Minor Categories   │          │
│        │  F.    Specify Single 5-Digit ICD9-CM Code   │          │
│        │  R.    Return to Main Options Menu           │          │
│        └──────────────────────────────────────────────┘          │
│                                                                   │
└─────────────────────────────────────────────────────────────────┘
```

FIGURE 5. ICD9-CM Selection Control Menu.

The following list describes the values extracted for each of the menu items (Figure 5):

A: all ICD9-CM categories lumped together,
B: each of the 29 major ICD9-CM categories separately,
C: each of the 130 minor ICD9-CM categories separately,
D: each user-specified major ICD9-CM category,
E: each user-specified minor ICD9-CM category,
F: a single ICD9-CM (NHRC Modified) code.

To specify major ICD9-CM categories, select "D. Specify Individual Major Categories" (Figure 5). The "Select Major ICD9-CM Categories" selection screen (Figure 6) will appear. Follow the instruction in section 4.1 to select major ICD9-CM categories. Exit when done. The program will return to the "Data Definition ICD9-CM Category" menu (Figure 4), display a message, and then return to the "New/Edit Data Structure" menu (Figure 4).

To specify minor ICD9 categories, select item "E. Specify Individual Minor Categories" (Figure 5). The "Select Minor ICD9 Categories" menu (Figure 7) will appear. Select a category from the list in Figure 6 and the appropriate category selection screen will appear.

As an example, select item "A. Viral Diseases" from Figure 7. The "Select Viral Diseases" selection screen (Figure 8) will appear. Follow the instruction in section 4.1 to select minor ICD9 viral disease categories. Note that the process (beginning with Figure 7) can be repeated as many times as necessary.

SELECT MAJOR ICD9 CATEGORIES

Viral Disease	Mental Disorders
Bacterial Diseases	Nervous System/Sense Organs
Mycobacterial Diseases	Circulatory System
Rickettsial Diseases	Respiratory System
Chlamydial Diseases	Digestive System
Sexually Transmitted Diseases	Genitourinary System
Fungus Diseases	Pregnancy/Childbirth
Spirochete Diseases	Skin/Subcutaneous Tissue
Protozoal Diseases	Musculoskeletal System
Ectoparasites	Congenital Anomalies
Worm-Caused Diseases	Perinatal Morbidity/Mortality
Diseases of Uncertain Etiology	Symtomatic/Ill-Defined
Neoplasms	Accident/Poisoning/Violence
Endocrine/Nutrition/Metabolic	Supplementary/Special Condition
Blood/Blood-Forming Organs	

Return to Previous Menu

FIGURE 6. Major ICD9 Category Selection Screen.

SELECT MINOR ICD9 CATEGORIES

A.   Viral Diseases
B.   Bacterial Diseases
C.   Mycobacterial Diseases
D.   Rickettsial Diseases
E.   Chlamydial Diseases
F.   Sexually Transmitted Diseases
G.   Fungus Diseases
H.   Spirochete Diseases
I.   Protozoal Diseases
J.   Ectoparasites
K.   Worm-Caused Diseases
L.   Diseases of Uncertain Etiology
M.   Other Diseases and Conditions
R.   Return to Previous Menu

FIGURE 7. First Screen in Selection Process for Minor ICD9 Categories.

```
+-----------------------------------------------------------+
|                                                           |
|  +-----------------------------------------------------+  |
|  |             SELECT VIRAL DISEASES                   |  |
|  +-----------------------------------------------------+  |
|  | Measles                  Tick-Borne Viral Encephalitis |
|  | Rubella                  Unsp. Arthropod-Borne Encephalitis |
|  | Chickenpox               Acute Encephalitis, Epidemic |
|  | Herpes zoster            Viral Encephlitis          |
|  | Herpes simplex           Arthropod-Borne Hemorrhagic Fever |
|  | Smallpox                 Epidemic Hemorrhagic Fever |
|  | Acute Poliomyelitis      Other Arthropod-BorneViral Diseases |
|  | Aseptic Meningitis (Enterovirus)  Viral Hepatitis   |
|  | Other Enterovirus Disease   Rabies                  |
|  | Cowpox                   Mumps                       |
|  | Other Viral Exanthemata  Coxsackie Virus            |
|  | Yellow Fever             Infectious Mononucleosis   |
|  | Dengue                   Other Viral Diseases, Conjunctiva |
|  | Mosquito-Borne Viral Encephalitis  Other Viral Diseases |
|  |                                                     |  |
|  |          Return to ICD9 Selection Menu              |  |
|  +-----------------------------------------------------+  |
|                                                           |
+-----------------------------------------------------------+
```

FIGURE 8. Minor ICD9 Category Selection Screen.

There are 13 category selection screens similar to Figure 8, one for each of the menu items "A" through "M" of Figure 7. After each exit from one of the 13 category selection screens, the "Select Minor ICD9 Categories" menu (Figure 7) will appear. Exit from Figure 7 when satisfied with the selections. The program will return to the "DATA DEFINITION - ICD9 CATEGORIES" menu (Figure 5).

To specify a 5-digit ICD-9 code, select "F. Specify Single 5-Digit ICD9 Code" (Figure 9). When the query appears, enter the desired code and press "ENTER." In this example, "30275" is entered.

```
+-----------------------------------------------------------+
|                                                           |
|  +-----------------------------------------------------+  |
|  |        DATA DEFINITION - ICD9 CATEGORIES            |  |
|  +-----------------------------------------------------+  |
|  | A.    Combine All Categories                        |  |
|  | B.    Specify All Major Categories                  |  |
|  | C.    Specify All Minor Categories                  |  |
|  | D.    Specify Individual Major Categories           |  |
|  | E.    Specify Individual Minor Categories           |  |
|  | F.    Specify Single 5-Digit ICD9 Code              |  |
|  | R.    Return to Previous Menu                       |  |
|  +-----------------------------------------------------+  |
|                                                           |
|          Enter 5-Digit ICD9 Code:  30275                  |
+-----------------------------------------------------------+
```

FIGURE 9. Entry Screen for 5-Digit ICD9 Code.

When satisfied with the ICD9 category selections, select "R. Return to Previous Menu" from the Data Definition - ICD9 Categories" menu (Figure 9). The program will return to the "New/Edit Data Structure" menu (Figure 4).

4.2.2 Specifying Sex Categories

Select "B. Sex Categories" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Sex" menu (Figure 10) will appear. All items will automatically make the appropriate entry in the data structure definition with no further user input.

```
DATA DEFINITION - SEX

A.      Combine Both Sexes
B.      Specify Male and Female
C.      Specify Male Only
D.      Specify Female Only
R.      Return to Previous Menu
```

FIGURE 10. Sex Category Selection Menu.

The following list describes the values extracted for each of the menu items (Figure 10):
A: both sexes lumped together,
B: male and female separately,
C: males only,
D: females only.

When satisfied with the sex category selections, select "R. Return to Previous Menu" from the "Data Definition - Sex" menu (Figure 10). The program will return to the "New/Edit Data Structure" menu (Figure 4).

4.2.3. Specifying Age Groups

Select "C. Age Groups" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Age Groups" menu (Figure 11) will appear. Items "A" and "B" will automatically make the appropriate entry in the data structure definition with no further user input. Item "C" requires further inputs.

```
DATA DEFINITION - AGE GROUPS

A.      Combine All Age Groups
B.      Specify Standard Age Groups
C.      Specify Nonstandard Age Groups
D.      Return to Previous Menu
```

FIGURE 11. Age Group Selection Control Menu.

-23-

The following list describes the values extracted for each of the menu items (Figure 11):

A: all ages are lumped together,

B: each of the following standard age groups: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-99,

C: each user-defined nonstandard age group.

To specify nonstandard age groups, select "C. Specify Nonstandard Age Groups" (Figure 11). The "Age Group Entry" screen (Figure 12) will appear. Enter the beginning age of each age group, then press "ENTER." The last entry must be the maximum age to be included for data extraction. Be sure to press "ENTER" after the last entry, then press "ESC."

AGE GROUP	BEGINNING AGE	ENDING AGE
1	18	
2	25	
3	30	
4	35	
5	40	
6	45	
7		
8		
.		
.		
.		
17		

Enter Starting Ages - Press <ESC> when done

FIGURE 12. Nonstandard Age Group Entry Screen.

After "ESC" is pressed, Figure 13 will appear showing the age groups chosen. If the selections are not correct, press "N" or "n." The program will return to Figure 12, display the previous entries, then allow changes (see section 4.1). After any changes are made, Figure 13 will appear again. The cycle can be repeated as many times as necessary. If the selections are correct (Figure 13), press "Y" or "y." The program will return to the "Data Definition - Age Groups" menu (Figure 11), display a message, and then return to the "New/Edit Data Structure" menu (Figure 4).

```
┌─────────────────────────────────────────────────────┐
│  ┌───────────────────────────────────────────────┐  │
│  │ AGE GROUP      BEGINNING AGE      ENDING AGE   │  │
│  │ ─────────────────────────────────────────────  │  │
│  │     1              18                 24        │  │
│  │     2              25                 29        │  │
│  │     3              30                 34        │  │
│  │     4              35                 39        │  │
│  │     5              40                 45        │  │
│  │     .                                           │  │
│  │     .                                           │  │
│  │     .                                           │  │
│  └───────────────────────────────────────────────┘  │
│                                                       │
│  ─────────────────────────────────────────────────   │
│          Are the Age Groups Correct? (Y/N):  Y        │
└─────────────────────────────────────────────────────┘
```

FIGURE 13. Nonstandard Age Group Verification Screen.

4.2.4 Specifying Service Branch Categories

Select "D. Service Branch Categories" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Service Branch" menu (Figure 14) will appear. All items will appear. All items will automatically make the appropriate entry in the data structure definition with no further user input.

```
┌─────────────────────────────────────────────────────┐
│  ┌───────────────────────────────────────────────┐  │
│  │      DATA DEFINITION - SERVICE BRANCH          │  │
│  │                                                 │  │
│  │      A.    Combine Both Branches                │  │
│  │      B.    Specify USN and USMC                 │  │
│  │      C.    Specify USN Only                     │  │
│  │      D.    Specify USMC Only                    │  │
│  │      R.    Return to Prɛvious Menu              │  │
│  └───────────────────────────────────────────────┘  │
└─────────────────────────────────────────────────────┘
```

FIGURE 14. Service Branch Selection Menu.

The following list describes the values extracted for each of the menu items (Figure 14):

A: both branches are lumped together,
B: USN and USMC separately,
C: USN only,
D: USMC only.

When satisfied with the service branch category selections, select "R. Return to Previous Menu" from the "Data Definition - Service Branch" menu (Figure 14). The program will return to the "New/Edit Data Structure" menu (Figure 4).

4.2.5 Specifying Platform Categories

Select "E. Platform Categories" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Platforms" menu (Figure 15) will appear. Items "A" through "D" will automatically make the appropriate entry in the data structure definition with no further user input. Item "E" requires further inputs.

```
DATA DEFINITION - PLATFORMS

A.    Combine All Platforms
B.    Specify Afloat and Ashore
C.    Specify All Ashore Platforms
D.    Specify All Afloat Platforms
E.    Specify Individual Platforms
R.    Return to Previous Menu
```

FIGURE 15. Platform Selection Control Menu.

The following list describes the values extracted for each of the menu items (Figure 15):

A: all platforms lumped together,
B: two groups: all ashore stations and all afloat platforms,
C: all ashore stations lumped together,
D: all afloat platforms lumped together,
E: each user-specified platform.

To specify major individual platforms, select "E. Specify Individual Platforms" (Figure 15). The "Select Platforms" selection screen will appear (Figure 16). Follow the instructions in section 4.1 to make selections, exit when done. The program will return to the "Data Definition - Platform" menu (Figure 15), display a message, and then return to the "New/Edit Data Structure" menu (Figure 4).

```
                SELECT PLATFORMS

        Ashore      AVT         LHD
        AE          BB          LKA
        AF          CA          LPH
        AFS         CG          LPD
        AGF         CCN         LSD
        AGS         CLG         LST
        AGSS        CV          MCM
        AO          CVA         MSO
        AOE         CVN         PG
        AOR         CVS         PHM
        AR          DD          SS
        ARL         DDG         SSBN
        ARS         FF          SSN
        AS          FFG         TAGS
        ASR         LCC         WMEC
        ATF         LHA

              Return to Previous Menu
```

FIGURE 16. Platform Category Selection Screen.

4.2.6 Specifying Oceanic Region Categories

Select "F. Oceanic Regions Categories" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Oceanic Regions" menu (Figure 17) will appear. Items "A" and "B" will automatically make the appropriate entry in the data structure definition with no further user input. Item "C" requires further inputs.
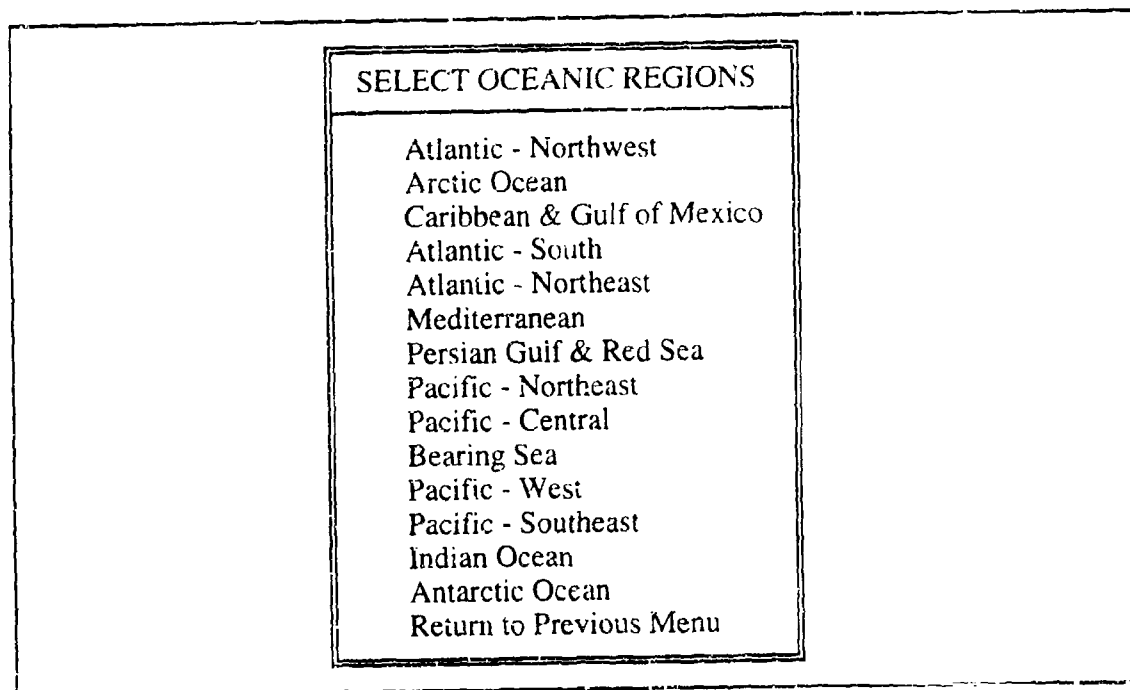
The following list describes the values extracted for each of the menu items (Figure 17):

        A: all regions lumped together,
        B: each region separately,
        C: each user-specified region.

```
      DATA DEFINITION - OCEANIC REGIONS

   A.    Combine All Oceanic Regions
   B.    Specify All Oceanic Regions
   C.    Specify Individual Oceanic Regions
   R.    Return to Previous Menu
```

FIGURE 17. Oceanic Region Selection Control Menu.

To specify individual oceanic regions, select "C. Specify Individual Oceanic Regions" (Figure 17). The "Select Oceanic Regions" selection screen will appear (Figure 18). Follow the instruction in section 4.1 to make selections.

```
┌─────────────────────────────────────────────────────┐
│        ┌───────────────────────────────────┐        │
│        │     SELECT OCEANIC REGIONS        │        │
│        ├───────────────────────────────────┤        │
│        │      Atlantic - Northwest         │        │
│        │      Arctic Ocean                 │        │
│        │      Caribbean & Gulf of Mexico   │        │
│        │      Atlantic - South             │        │
│        │      Atlantic - Northeast         │        │
│        │      Mediterranean                │        │
│        │      Persian Gulf & Red Sea       │        │
│        │      Pacific - Northeast          │        │
│        │      Pacific - Central            │        │
│        │      Bearing Sea                  │        │
│        │      Pacific - West               │        │
│        │      Pacific - Southeast          │        │
│        │      Indian Ocean                 │        │
│        │      Antarctic Ocean              │        │
│        │      Return to Previous Menu      │        │
│        └───────────────────────────────────┘        │
└─────────────────────────────────────────────────────┘
```

FIGURE 18. Oceanic Region Category Selection Screen.

Exit when done. The program will return to the "Data Definition - Oceanic Regions" menu (Figure 17), display a message, and then return to the "New/Edit Data Structure" menu (Figure 4).

4.2.7 Specifying Rate/Occupation Code

To specify a rate/occupation code, select "G. Specify Rate/Occupation Code" (Figure 19). A query will appear under the menu as shown. Enter the desired code and press "ENTER"; for example, enter "10." If a code is entered, only data for that code will be extracted. If no code is entered, rate/occupation code is ignored during data extraction. After the entry is made, the program will display a message and then return to the "New/Edit Data Structure" menu (Figure 4).

```
                    NEW/EDIT DATA STRUCTURE

          A.      ICD9 Categories
          B.      Sex Categories
          C.      Age Groups
          D.      Service Branch Categories
          E.      Platform Categories
          F.      Oceanic Region Categories
          G.      Rate/Occupation Code
          H.      Inclusive Dates
          I.      Calculation Parameters
          J.      Cancel All Changes
          R.      Return to Previous Menu
```

FIGURE 19. Rate/Occupation Code Data Entry Screen.

4.2.8 Specifying Inclusion Dates

Select "H. Specify Inclusion Dates" (Figure 20). A query will appear under the menu as shown. Enter the desired dates, pressing "ENTER" after each; for example enter "01/01/80" and "12/31/85." If dates are entered, only data between and including those dates will be extracted. If no dates are entered, dates will be ignored during data extraction. After the entries are made, the program will display a message and then return to the "New/Edit Data Structure" menu (Figure 4).

```
                    NEW/EDIT DATA STRUCTURE

          A.      ICD9 Categories
          B.      Sex Categories
          C.      Age Groups
          D.      Service Branch Categories
          E.      Platform Categories
          F.      Oceanic Region Categories
          G.      Rate/Occupation Code
          H.      Inclusive Dates
          I.      Calculation Parameters
          J.      Cancel All Changes
          R.      Return to Previous Menu


          Enter Starting Date:          01/01/80
          Enter Ending Date:            12/31/85
```

FIGURE 20. Inclusion Date Entry Screen.

4.2.9 Specifying Calculation Parameters

Select "I. Calculation Parameters" from the "New/Edit Data Structure" menu (Figure 4). The "Data Definition - Calculations" menu (Figure 21) will appear. All items will automatically make the appropriate entry in the data structure definition with no further user input.
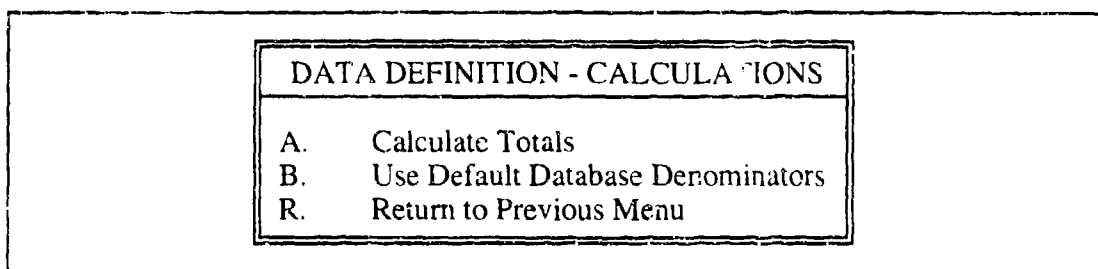
```
┌──────────────────────────────────────────────────┐
│    ┌──────────────────────────────────────────┐  │
│    │  DATA DEFINITION - CALCULA ˥IONS         │  │
│    │                                          │  │
│    │  A.    Calculate Totals                  │  │
│    │  B.    Use Default Database Denominators │  │
│    │  R.    Return to Previous Menu           │  │
│    └──────────────────────────────────────────┘  │
└──────────────────────────────────────────────────┘
```

FIGURE 21. Calculation Parameter Selection Menu.

The following list describes values extracted for each of the menu items (Figure 21):

A: total number of incidences,
B: rates normalized to predefined denominators.

Item "A" is the default; it will be used if item "B" is not chosen explicitly. When satisfied with the calculation parameter selection, select "R. Return to Previous Menu" (Figure 21). The program will return to the "New/Edit Data Structure" menu (Figure 4).

4.2.10 Canceling New Definition

Select "J. Cancel All Changes" (Figure 22). A warning message will appear as shown. To cancel all changes, press "Y" or "y." To continue without canceling changes, press "N" or "n." In either case, the program will return to the "New/Edit Data Structure" menu (Figure 4).

4.2.11 Saving the New Data Structure Definition

After all selections have been made for the new data structure definition, the program will be in the "New/Edit Data Structure" menu (Figure 4). Select "R. Return to Previous Menu." The "Data Structure Control" menu (Figure 23) will appear.

```
  ┌──────────────────────────────────────┐
  │       NEW/EDIT DATA STRUCTURE        │
  ├──────────────────────────────────────┤
  │  A.    ICD9 Categories               │
  │  B.    Sex Categories                │
  │  C.    Age Groups                    │
  │  D.    Service Branch Categories     │
  │  E.    Platform Categories           │
  │  F.    Oceanic Region Categories     │
  │  G.    Rate/Occupation Code          │
  │  H.    Inclusive Dates               │
  │  I.    Calculation Parameters        │
  │  J.    Cancel All Changes            │
  │  R.    Return to Previous Menu       │
  └──────────────────────────────────────┘

  WARNING - This Option Will Destroy All
  Active Data Strucute Changes Do You Wish to
  Continue? (Y/N): N
```

FIGURE 22. Warning Screen for Cancellation of All New Inputs.

Select "E. Save Active Definition." The first query shown in Figure 23 will appear. Enter the name to be given to the new data stucture definition and press "ENTER." The name must be alphanumeric, without spaces, and between one and eight characters long. If a previous data structure definition has been given the name entered, the second query shown in Figure 23 will appear. After the two queries are answered, the program will return to the "New/Edit Data Structure" menu (Figure 4). Note that in order to review or print a data structure definition, it must first have been saved.

```
  ┌──────────────────────────────────────┐
  │        DATA STRUCTURE CONTROL        │
  ├──────────────────────────────────────┤
  │  A.    Create/Edit Definition        │
  │  B.    Select Existing Definition    │
  │  C.    Review Active Definition      │
  │  D.    Print Active Definition       │
  │  E.    Save Active Definition        │
  │  R.    Return to Main Options Menu   │
  └──────────────────────────────────────┘

  Enter Name of Data Definition to Save:  TEST1
      Definition Exists.  Overwrite (Y/N):  N
```

FIGURE 23. Save Data Structure Definition Query Screen.

4.3 Retrieving and Editing a Data Structure Definition

This section describes the techniques for editing an existing data structure definition. The techniques are, for the most part, identical to those of section 4.2 for entering a new definition (with the exceptions noted). For the basic techniques, refer to section 4.2.

4.3.1 Editing an Existing Data Structure Definition

Epibase automatically detects when information being entered is new or when it is part of an existing data structure definition. A existing definition is edited using the same menus and techniques as for entering a new definition (see section 4.2) . The only difference is that previous entered selections will appear when menus, data entry screens, and category selection screens are selected. Any entry can be changed at any time.

Changes do not become permanent until the edited data structure definition has been saved (section 4.2.10). Therefore, changes must be saved before they can be reviewed or printed. If a definition is edited and then saved under a name other than its original name, the original definition still exists under the old name. The new name becomes the name of the active definition.

4.3.2 Retrieving an Existing Data Structure Definition

From the "Main Options Menu (Figure 2) select "A. Define Data Structure." The "Data Structure Control" menu will appear (Figure 24).

Select "B. Select Existing Definition" and the query shown will appear. Enter the name of a previously saved data structure definition and press "ENTER." The name must be alphanumeric, without spaces, and between one and eight characters long. For example, enter "TEST1."

```
              DATA STRUCTURE CONTROL

    A.      Create/Edit Definition
    B.      Select Existing Definition
    C.      Review Active Definition
    D.      Print Active Definition
    E.      Save Active Definition
    R.      Return to Main Options Menu
```

FIGURE 24. Existing Definition Selection Query Screen.

If the definition name does not exit, the warning message will appear. After reading the message, press any key and the program will return to the "Data Structure Control" menu. If the definition name does exist, the definition is read into the program automatically and becomes the active definition, and is available for editing or other operations.

Two utility funtions are useful for developing new data structure definitions or editing old ones. These functions are used to review the active definition on the screen and to print the active definition.

4.4.1 Reviewing the Active Definition

To review the active data structure definition, select "C. Review Active Definition" from the "Data Structure Control" menu (Figure 25). If no definition has been made active or if a new definition has not been previously saved, the error message shown will appear. After reading the message, press any key to return to the "Data Structure Control" menu.

If the data structure definition is currently active, the definition will appear on the screen and can be scrolied and viewed using the ARROW, PGUP. and PGDN keys. Figure 26 is an example of a data structure definition.

```
        ┌──────────────────────────────────────┐
        │       DATA STRUCTURE CONTROL         │
        ├──────────────────────────────────────┤
        │                                      │
        │   A.     Create/Edit Definition      │
        │   B.     Select Existing Definition  │
        │   C.     Review Active Definition    │
        │   D.     Print Active Definition     │
        │   E.     Save Active Definition      │
        │   R.     Return to Main Options Menu │
        └──────────────────────────────────────┘

      Definition Must Have Been Saved Before It Can Be Viewed
                    Press Any Key to Continue
```

FIGURE 25. Review Active Definition Selection and Warning Screen.

```
DATA STRUCTURE DEFINITION - TEST1
─────────────────────────────────────────────────
ICD9  CATEGORIES:       (9)
Code  description
─────────────────────────────────────────────────
  2    Rubella
  4    Herpes zoster
  5    Herpes simplex
  6    Cowpox
  7    Mosquito-Borne Viral Encephalitis
  8    Viral Hepatitis
  9    Other Viral Diseases
```

FIGURE 26. Example Data Structure Definition

SEX:		
Code	description	
	2 Male and Female Separately	

ACE GROUPS:		(4)
Start	End	
18	24	
25	29	
30	34	
35	40	

SERVICE BRANCES:	
Code	description
3 USN Only	

PLATFORMS:		(5)
Code	description	
1	Ashore	
4	AFS	
27	DD	
28	DDG	
36	LPH	

OCEAN REGIONS:		(2)
Code	description	
3	Caribbean Sea and Gulf of Mexico	
12	Pacific Ocean - Southwest	

RATE/OCCUPATION:	
Code	description
10 Machinist	

CALCULATION PARAMETERS:	
Code	description
1 Calculate Totals Only	

INCLUSIVE DATES:	
Start	Stop
01/01/80	12/31/85

FIGURE 26. (continued)

4.4.2 Printing the Active Definition

To print the active data structure definition, select "D. Print Active Definition" from the "Data Structure Control" menu (Figure 27). If no definition has been made active or if a new definition has not been previously saved, the error message shown will appear. After reading the message, press any key to return to the "Data Structure Control" menu. If the data structure definition is currently active, the definition will be printed. The printout will appear as in Figure 26.
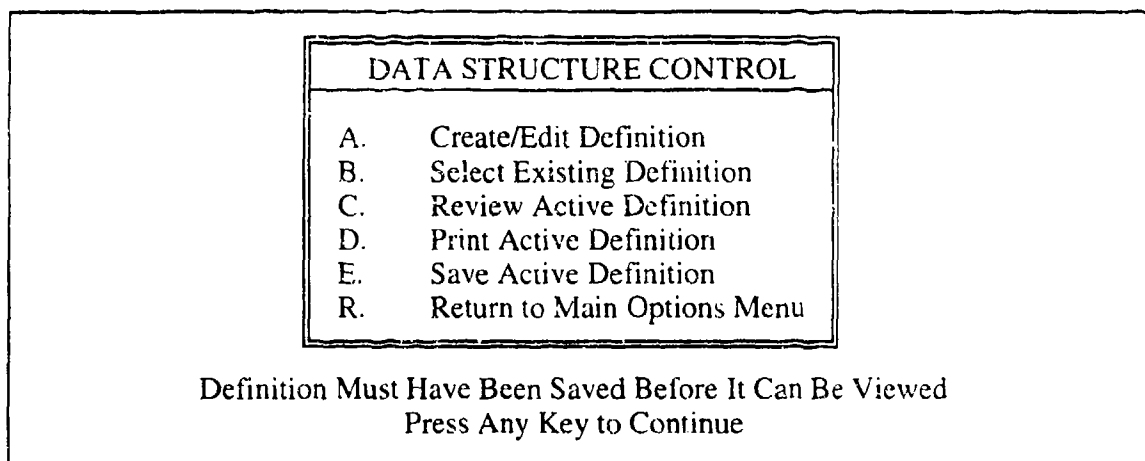
```
┌─────────────────────────────────────────────────────────┐
│         ┌─────────────────────────────────────┐         │
│         │       DATA STRUCTURE CONTROL        │         │
│         ├─────────────────────────────────────┤         │
│         │                                     │         │
│         │   A.    Create/Edit Definition      │         │
│         │   B.    Select Existing Definition  │         │
│         │   C.    Review Active Definition    │         │
│         │   D.    Print Active Definition     │         │
│         │   E.    Save Active Definition      │         │
│         │   R.    Return to Main Options Menu │         │
│         └─────────────────────────────────────┘         │
│                                                         │
│     Definition Must Have Been Saved Before It Can Be Viewed │
│                 Press Any Key to Continue               │
└─────────────────────────────────────────────────────────┘
```

FIGURE 27. Review Active Definition Selection and Warning Screen.

4.5 Data Extraction

In this section, the method of extracting data according to a data structure definition created and saved using the methods of section 4.2 through 4.4 is presented. As before, when Epibase is invoked, the "Main Options Menu" appears (Figure 28).

```
┌─────────────────────────────────────────────────────────┐
│         ┌─────────────────────────────────────┐         │
│         │          MAIN OPTIONS MENU          │         │
│         ├─────────────────────────────────────┤         │
│         │                                     │         │
│         │   A.    Define Data Structure       │         │
│         │   B.    Extract Data                │         │
│         │   Q.    Exit Program                │         │
│         └─────────────────────────────────────┘         │
└─────────────────────────────────────────────────────────┘
```
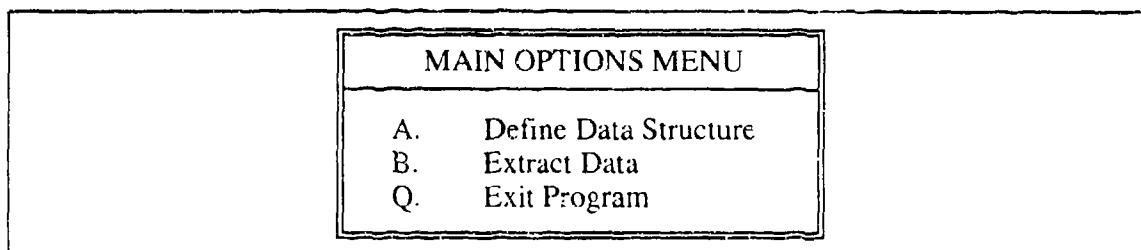
FIGURE 28. Main Options Menu.

Select "B Extract Data." The "Data Extraction" menu (Figure 29) will appear. Select "A. Perform Data Extraction" and the "Enter Name" query will appear. Enter the name of a previously saved data structure definition and press "ENTER." The name must be alphanumeric, without spaces, and between one and eight characters long. For example, enter "TEST1."

```
┌─────────────────────────────────────────────────────┐
│        ┌───────────────────────────────────┐        │
│        │          DATA EXTRACTION          │        │
│        │───────────────────────────────────│        │
│        │                                   │        │
│        │  A.    Perform Data Extraction    │        │
│        │  B.    Review Data on Screen      │        │
│        │  C.    Output Data to Printer     │        │
│        │  R.    Return to Main Options Menu│        │
│        │                                   │        │
│        └───────────────────────────────────┘        │
│                                                     │
│     Enter Name of Data Definition to Make Active:  TEST1 │
│                                                     │
│          Extracting Data - Please Wait              │
│          Definition Does Not Exist                  │
│          Data Extraction Complete                   │
│          Press Any Key to Continue                  │
│                                                     │
└─────────────────────────────────────────────────────┘
```

FIGURE 29. Data Extraction Query and Message Screen.

If the data structure definition name does not exist, the warning message will appear. Press any key to return to the "Data Extraction" Menu. If the name does exist, the definition will be read into the program, made the active definition, and data extraction will begin. While data is being extracted, the "Extracting Data - Please Wait" message will remain on the screen. When data extraction is complete, the "Data Extraction Complete" message will appear. When the latter appears, press any key to return to the :Data Extraction" menu.

4.6 Data Extraction Utility Functions

There are two data extraction utility functions. One is used to view and the other to print the extracted data. Both functions are discussed in this section. The output of both functions are raw ASCII data files containing the extracted data.

4.6.1 Reviewing Extracted Data

Select "B. Review Data on Screen" from the "Data Extraction" menu (Figure 30). If no data has been extracted based on the current active data structure definition, the error message will appear. After reading the message, press any key to return to the "Data Extraction" menu.

If data has been extracted based on the current active data structure definition, the data will appear on the screen in a form that can be scrolled and viewed using the ARROW, PGUP, and PGDN keys.

```
                    DATA EXTRACTION

        A.      Perform Data Extraction
        B.      Review Data on Screen
        C.      Output Data to Printer
        R.      Return to Main Options Menu

      Data Must Have Been Extracted Before It Can Be Viewed

                  Press Any Key to Continue
```

FIGURE 30. Data Review Selection Screen.

4.6.2 Printing Extracted Data

Select "C. Output Data to Printer" from the "Data Extraction" menu (Figure 31). If no data has been extracted based on the current active data structure definition, the error message will appear. If a hardware error occurs sending the data to the printer, the "Printer Not Ready" message will appear. After reading the message, press any key to return to the "Data Extraction" menu.

If data has been extracted based on the current active data stucture definition and the printer is ready, the data will be printed.

```
                    DATA EXTRACTION

        A.      Perform Data Extraction
        B.      Review Data on Screen
        C.      Output Data to Printer
        R.      Return to Main Options Menu

      Data Must Have Been Extracted Before It Can Be Viewed

                    Printer Not Ready
                  Press Any Key to Continue
```

FIGURE 31. Data Print Selection Screen.

5.0 EPILIMIT: ICD9-CM RATE SURVEILLANCE

This section describes the Epilimit rate surveillance module. Epilimit has three functions in all: automatic and on-demand rate surveillance and quality control. Automatic rate surveillance and quality control are invoked by EPISYS without user intervention. On-demand rate surveillance is called by the user, but is automatic in its operation after it is invoked.

5.1 Automatic and On-Demand Rate Surveillance

Automatic rate surveillance is completely transparent to the user. On-demand rate surveillance is called by the user and then the remainder of its operation is completely transparent. This section sets forth the statistical techniques for both processes.

It was found during the development and testing of Episam that almost all of the 130 minor ICD9-CM categories showed a periodicity with two strong peaks of less that one week. Upon further investigation, these peaks were found to be related to the operation schedule of USN and USMC clinics. Therefore, the lower resolution limit for the detection of ICD9 rate trends over time is taken as one week and all Epilimit procedures operate on weekly averaged rates.

Each of the 130 minor ICD9 categories is analyzed separately and independently from one another. The following dependent variables are analyzed for each minor ICD9 category: (1) total number of incidences, and (2) normalized rates. In addition, both dependent variables are analyzed within each cell of a standard multidimensional data struction that differs between automated and on-demand surveillance. Automated surveillance uses a data structure defined by service branch, pay grade, assignment, and ocean region. On-demand surveillance uses a data structure created by Epibase; it is based on the independent variables used by the automated process plus sex, age group, and inclusion dates.

The following discussion shows the calculations for a single analytical cell. For total number of incidences, the dependent variables is

$$x_i = \frac{1}{7} \sum_{j \cdot s_i}^{s_i + 6} N_j \tag{1}$$

where i = 1, 2, ... (week since the beginning of the data set), N_j is the total for day j, and s_i is the starting day for week i.

For normalized rates

$$x_i = \frac{1}{7} \sum_{j-s_1}^{s_1+6} \frac{N_j}{D_i} \tag{2}$$

where D_i is the appropriate denominator for the data structure cell and week. From this point, both dependent variables are treated alike and are henceforth referred to simply as the dependent variable.

The analysis uses a simple regression technique to detect trends in the dependent variable. The regression model is

$$Y_i = b_o + b_1 Z_i + b_2 Z_i^2 \tag{3}$$

where

$$Z_i = \sum_{j-1}^{i} (x_i - \mu) \tag{4}$$

and μ is the predetermined (expected) rate or total value that is the basis of the analysis and $i = 1, 2, \ldots, n$ where n is the number of weeks in the analysis.

Note that Z_i is a cumulative sum of "i" deviations since the beginning of the analysis. If no trends occur in the analysis, $X_i - \mu$ is a random error with zero mean and Y_i (equation 4) is a sum of the errors, also with zero mean. If, however, X_i shows a trend away from μ beginning at some week "k", then

$$Z_i = \sum_{j-1}^{i} (x_i - \mu) + (n - k)\delta \tag{5}$$

where δ is the mean deviation. There is now the original random element with zero mean (equation 4) plus a systematic component that increases with each new surveillance week. Any trend will thus be indicated by a change in slope in equation 3 (b_1 and $b_2 = 0$). Note that b_0 represents the "zero intercept" and is assumed to be always zero.

As each new week is added, equation 3 is estimated and the coefficients (b_1 and b_2) are tested for deviation from zero. To illustrate this procedure, let Z be the matrix of values derived from equation 4 (Z and Z^2) and let Y be the vector of values from equation 3. In matrix terms, the coefficients are estimated as

$$b_i = (Z'Z)^{-1} (Z'Y).\qquad(6)$$

Now the sums of squares total (SST), regression (SSR), and error (SSE) are

$$SST = Y'Y$$
$$SSR = b_i (Z'Y)\qquad(7)$$
$$SSE = SST - SSR$$

and the variance as

$$S^2 = \frac{SSE}{i-3}.\qquad(8)$$

Finally the test statistic is calculated as

$$T = \frac{b_i}{\sqrt{S^2 C_{ii}}}\qquad(9)$$

where

$$C_{ii} = (Z'Z)^{-1}_{ii}\qquad(10)$$

and T is distributed as Student's T with i-3 degrees of freedom. If $Pr\ (T \neq 0) > 1 - \alpha$, then a significant trend has been detected and is reported.

5.2 Quality Control

Quality control involves two steps: (1) an automated procedure to detect possible prob em areas, and (2) possible hand-checking if serious problems are detected. In the automated procedure, each encounter is defined as a data record. Each variable is termed a data element. As records are manipulated by Epibase, they are error checked. Each data element that is out of range is counted as one data error; thus there can be more than one data error per record. Each required data element that is missing is also counted as a data error.

The probability of an error in a record is defined as

$$p = \frac{\sum\limits_{i=1}^{N} a_i}{\sum\limits_{i=1}^{N} m_i} \qquad (11)$$

where data records are arranged in batches of size "N," the number of data elements in record "i" is defined as "m_i," and the number of data errors in record "i" is defined as a_i. With these definitions, the following procedure is an exact test for $Ho: p \leq e$, where e is a predefined maximum allowable error rate.

Batch records in groups of size "N." Randomly choose "n" records from the batch and obtain m_i and a_i for each of the "n" records. When choosing a value for "n," choose the smallest "n" such that

$$e \sum\limits_{i=1}^{n} m_i \geq 10. \qquad (12)$$

Now compute "p" from equation 11, substituting "n" for "N" in the summation. Then calculate the following mean and variance

$$m = \frac{1}{n} \sum\limits_{i=i}^{n} m_i \qquad (13)$$

$$Var(n_p) = \frac{n(1 - \frac{n}{N})}{(n-1)m^2} \cdot \sum\limits_{i=1}^{n} a_i^2 - 2p \sum\limits_{i=1}^{n} a_i \, m_i + p^2 \sum\limits_{i=1}^{n} m_i^2 \qquad (14)$$

So long as the criterion of equation 12 is met, the following standard normal test statistic can be calculated:

$$T = \frac{n_p - ne}{\sqrt{\dfrac{Var(n_p)}{n}}} \qquad (1.)$$

The null hypothesis $Ho: p \leq e$ is rejected if $Pr(T > 0) > 1 - \alpha$, indicating that the maximum allowable error rate has been exceeded; T is distributed as a standard normal with $"n-1"$ degrees of freedom. If the null hypothesis is rejected, the entire batch of $"N"$ records must be carefully hand-checked and corrections made.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE July 1992	3. REPORT TYPE AND DATE COVERED Software Documentation

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Development of an Epidemiological Data Base Management, Extraction, and Analysis System (EPISYS)	Program Element: 63706N Work Unit Number:
6. AUTHOR(S) Ivan Show, Ph.D. Martin R. White, M.P.H.	M0095.005-6103

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION
Naval Health Research Center P. O. Box 85122 San Diego, CA 92186-5122	Report No. Technical Document No.93-6F

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Naval Medical Research and Development Command National Naval Medical Center Building 1, Tower 2 Bethesda, MD 20889-5044	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited.	

13. ABSTRACT *(Maximum 200 words)*

This report is the user documentation for the integrated system of software programs called Epidemiological Interactive System (EPISYS). The program consists of a number of user-callable modules, Epilimit, Epibase, Episaam, Epimips, and a Utilities module. The system, when complete, will give researchers and medical planners the capability to easily detect, sample, and analyze any ICD9-CM illness based upon a number of ancillary variables including age, race, sex, service branch, ship type, paygrade, and occupation. These programs will significantly improve the ability to access our Medical History Files and will provide investigators with an integrated system of computer programs for health monitoring and medical projection needs. The user documentation in this report explains in detail the user modules, Epibase, and Epilimit.

14. SUBJECT TERMS		15. NUMBER OF PAGES 42
Epidemiology, Occupational Safety and Health, Medical Management Tools, and Medical Data base Software		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited